



Aldridge, M. P., Johnson, O. T., & Scarlett, J. (2016). Improved group testing rates with constant column weight designs. In *2016 IEEE International Symposium on Information Theory (ISIT 2016) : Proceedings of a meeting held 10-15 July 2016 at Barcelona, Spain* (pp. 1381-1386). (Proceedings of the IEEE International Symposium on Information Theory (ISIT)). Institute of Electrical and Electronics Engineers (IEEE). <https://doi.org/10.1109/ISIT.2016.7541525>

Peer reviewed version

Link to published version (if available):  
[10.1109/ISIT.2016.7541525](https://doi.org/10.1109/ISIT.2016.7541525)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Institute of Electrical and Electronics Engineers at <http://dx.doi.org/10.1109/ISIT.2016.7541525>.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Improved Group Testing Rates with Constant Column Weight Designs

Matthew Aldridge

Department of Mathematical Sciences  
University of Bath  
Bath, BA2 7AY, UK  
Email: m.aldridge@bath.ac.uk

Oliver Johnson

School of Mathematics  
University of Bristol  
Bristol, BS8 1TW, UK  
Email: maotj@bristol.ac.uk

Jonathan Scarlett

Laboratory for Information and Inference Systems  
École Polytechnique Fédérale de Lausanne  
CH-1015, Switzerland  
Email: jonathan.scarlett@epfl.ch

**Abstract**—We consider nonadaptive group testing where each item is placed in a constant number of tests. The tests are chosen uniformly at random with replacement, so the testing matrix has (almost) constant column weights. We show that performance is improved compared to Bernoulli designs, where each item is placed in each test independently with a fixed probability. In particular, we show that the rate of the practical COMP detection algorithm is increased by 31% in all sparsity regimes. In dense cases, this beats the best possible algorithm with Bernoulli tests, and in sparse cases is the best proven performance of any practical algorithm. We also give an algorithm-independent upper bound for the constant column weight case; for dense cases this is again a 31% increase over the analogous Bernoulli result.

## I. INTRODUCTION

The group testing problem, introduced by Dorfman [7], is described as follows. Suppose we have a number of items, some of which are defective, and carry out a series of tests on subsets of items (‘pools’). In the standard noiseless model we consider here, the result of a test is positive if the pool contains at least one defective item, and is negative otherwise. The task is to detect which items are defective using as few tests as possible, using only the list of testing pools and the outcomes of the corresponding tests.

We let  $N$  denote the total number of items, let  $K$  denote the number of defective items, and focus on the regime  $K = o(N)$ . We suppose that the set  $\mathcal{K}$  of defective items is chosen uniformly at random from the  $\binom{N}{K}$  possible defective sets.

Finding the true defective set requires us to learn  $\log_2 \binom{N}{K}$  bits of information. If an algorithm uses  $T$  tests, then, following [5], we can consider the number of bits of information learned per test  $\log_2 \binom{N}{K}/T$  as the *rate* of the algorithm, and consider the *capacity* as the supremum of all rates that can be achieved by any algorithm. We consider performance in the regime where the number of defectives scales as  $K = \Theta(N^\theta)$  for some density parameter  $\theta \in (0, 1)$ .

In the adaptive case – where we choose successive testing pools using the outcomes of previous tests – Hwang’s generalized binary splitting algorithm [9] shows that the capacity is 1 for all  $\theta \in (0, 1)$  [5]. In this paper, we consider the nonadaptive case, where the testing pools are chosen in advance.

It will be useful to list the testing pools in a binary matrix  $X \in \{0, 1\}^{T \times N}$ , where  $x_{ti} = 1$  denotes that item  $i$  is included

in test  $t$ , with  $x_{ti} = 0$  otherwise. Hence, the rows of the matrix correspond to tests, and the columns correspond to items.

We observe the outcomes  $\mathbf{y} = (y_t) \in \{0, 1\}^T$ . A positive outcome  $y_t = 1$  occurs if there exists a defective item in that test; that is, if for some  $i \in \mathcal{K}$  we have  $x_{ti} = 1$ . A negative outcome  $y_t = 0$  occurs otherwise.

As described in [8], it appears to be difficult to design a matrix with order-optimal performance using combinatorial constructions. Hence, a great deal of recent work on non-adaptive group testing has considered random design matrices, with a particular focus on Bernoulli random designs, in which each item is placed in each test independently with a given probability; see for example [1]–[4], [6], [14], [15]. The capacity for Bernoulli nonadaptive testing is

$$C(\theta) = \max_{\nu > 0} \min \left\{ \frac{\nu e^{-\nu}}{\ln 2} \frac{1 - \theta}{\theta}, h(e^{-\nu}) \right\}, \quad (1)$$

in particular yielding a capacity of 1 for  $\theta \leq 1/3$ . The achievability part of this result was given by Scarlett and Cevher [14], and the converse by Aldridge [1].

This paper shows that improvements are possible with a different class of test designs, where each item is placed in a fixed number  $L$  of tests, with the tests chosen uniformly at random with replacement. That is, independently within each column of  $X$ ,  $L$  random entries are selected uniformly at random with replacement and set to 1. We refer to these as ‘constant column weight’ designs – although strictly speaking, since the sampling is with replacement, some columns will have weight slightly less than  $L$ . The results of this paper could also be derived by sampling without replacement, thus yielding strictly constant column weights, with each item in *exactly*  $L$  tests, but we find that considering replacement is more convenient and leads to shorter proofs. It will be convenient to parametrise  $L = \nu T/K$ , and we will later see that  $\nu = \ln 2 \approx 0.693$  optimises our bounds for all  $\theta$ . This choice of  $\nu$  gives a 50 : 50 chance of tests being positive or negative, in contrast to Bernoulli testing with  $\theta > 1/3$ , where the optimal procedure has on average more positive than negative tests.

The main result of this paper (Theorem 1) shows that with a constant column weight design, a simple and practical algorithm called COMP [6] achieves a rate of  $0.693(1 - \theta)$ .

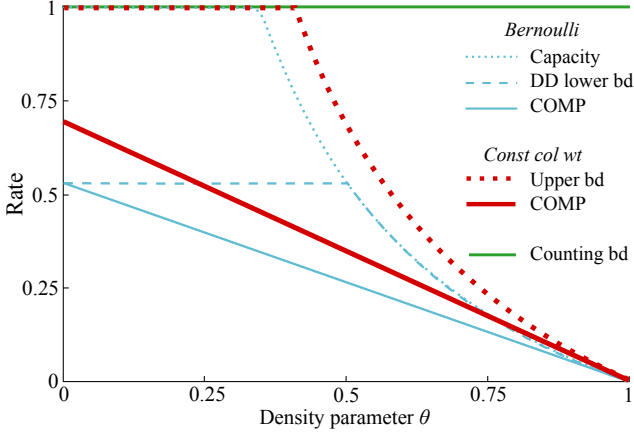


Fig. 1. Graph showing rates and bounds for group testing algorithms with Bernoulli designs and constant column weight designs.

When used with a Bernoulli design, COMP has maximum rate  $0.531(1 - \theta)$ , so we achieve a rate increase of 30.6% for all  $\theta$ . Further, COMP with a constant column weight design outperforms *any* algorithm used with a Bernoulli design for  $\theta > 0.766$ , and gives the best proven rate for a practical algorithm for  $\theta < 0.234$  (beating a bound on Bernoulli designs with the DD algorithm [3]). These rates are shown in Fig. 1.

In addition, we provide an algorithm-independent converse (Theorem 2); that is, an upper bound on the rate that can be achieved by any detection algorithm with a constant column weight design. We conjecture that, as in [1], this converse is sharp and is achieved by the SSS algorithm.

We also give empirical evidence that, for a variety of algorithms, constant column weight designs improve on Bernoulli designs in the finite blocklength regime (see Fig. 2).

The idea of using constant column weight matrix designs is not a new one. The key contribution of this paper is to rigorously analyse the performance of such designs, and to show that they can out-perform Bernoulli designs. We briefly mention some previous works that used constant column weight (and other related) designs.

Mézard *et al.* [13], consider randomized designs with both fixed row and column weights, and with fixed column weights only. It is noted therein that such designs can beat Bernoulli designs; however, they note that the analysis relies on a ‘short-loops’ assumption that is shown to be rigorous only for  $\theta > 5/6$ , and in fact shown to be invalid for small  $\theta$ . In contrast, we present analysis techniques that are rigorous for all  $\theta \in (0, 1)$ .

Kautz and Singleton [11] observe that matrices corresponding to constant weight codes with a high minimum distance perform well in group testing. However, controlling the minimum distance is known to be a stringent requirement.

Wadayama [16] analyses constant row and column weight designs in the  $K = cN$  regime, and demonstrates close-to-optimal asymptotic performance for certain ratios of parameter sizes.

Chan *et al.* [6] consider designs with constant row-weights only (like here, sampling with replacement) and find no improvement over Bernoulli designs.

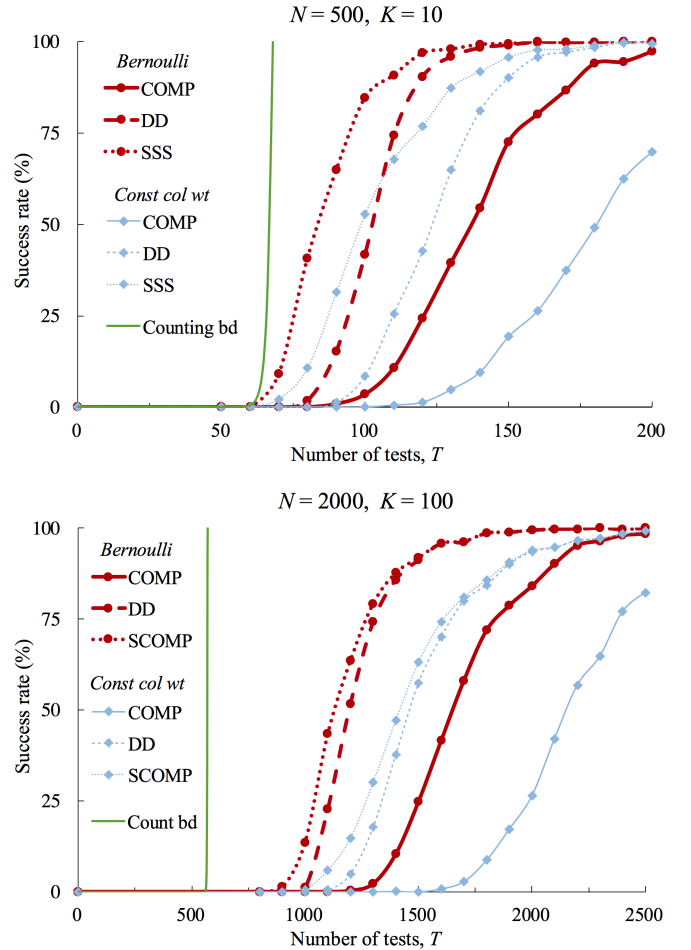


Fig. 2. Empirical performance results in the cases  $N = 500$ ,  $K = 10$ , and  $N = 2000$ ,  $K = 100$ , for a variety of algorithms with constant column weight and Bernoulli designs. Each point represents 1000 experiments.

The rest of this paper is organised as follows. Section II demonstrates the empirical performance of the designs and algorithms discussed in this paper. Section III introduces some necessary results on the classical ‘coupon collector’ problem. Section IV defines the COMP algorithm, and proves our main theorem on the maximum rate for COMP with constant column weight designs (Theorem 1). Section V defines the SSS algorithm and proves an algorithm-independent upper bound for constant column weight designs.

## II. SIMULATIONS

Fig. 2 shows empirical performance results in an illustrative smaller, sparser case ( $N = 500$ ,  $K = 10 = N^{0.371}$ ), and an illustrative larger denser case ( $N = 2000$ ,  $K = 100 = N^{0.606}$ ). In the first case, we show the COMP and SSS algorithms studied here, and the practical Definite Defectives (DD) algorithm [3]. In the second case, SSS is impractical, so we use the Sequential COMP (SCOMP) approximation [3]. Note that all algorithms perform better under constant column weight designs, and the simple DD algorithm with a constant column weight design usually outperforms more complicated algorithms with a Bernoulli design. We also plot the ‘counting

bound' universal converse (see [5], [10]), which shows that for any design (adaptive or nonadaptive), any algorithm has a success probability bounded above by  $2^T / \binom{N}{K}$ . In the first case of Fig. 2 the constant column weight SSS algorithm has empirical performance close to this universal upper bound.

### III. COUPON COLLECTOR RESULTS

Recall the coupon collector problem, where uniformly random selections, with replacement, are made from a population of  $T$  different coupons. After making  $c$  selections, the collector has collected some (random) number  $W$  of *distinct* coupons. Clearly  $W$  can be as large as  $c$  if all  $c$  selections are different, but it can be less if there are some 'repeated' coupons. The following results will be useful later.

*Lemma 1:* Consider a population of  $T$  coupons. A collector makes  $c$  selections, and finds she has  $W(c)$  distinct coupons.

1) We have

$$\mathbb{E}W(c) = \left(1 - \left(1 - \frac{1}{T}\right)^c\right) T, \quad (2)$$

and further, if  $c = \alpha T$  for some  $\alpha \in (0, 1)$ , then

$$\mathbb{E}W(\alpha T) \sim (1 - e^{-\alpha})T$$

as  $T \rightarrow \infty$  (where here and subsequently  $\sim$  denotes equality up to a multiplicative  $1 + o(1)$  term).

2) Again when  $c = \alpha T$ , we have concentration of  $W$  about its mean, in that for any  $\epsilon > 0$ ,

$$\mathbb{P}(|W(\alpha T) - (1 - e^{-\alpha})T| \geq \epsilon T) \leq 2 \exp\left(-\frac{\epsilon^2 T}{\alpha}\right),$$

for  $T$  sufficiently large.

*Proof:* For part 1, by linearity of expectation, we have

$$\begin{aligned} \mathbb{E}W(c) &= \sum_{j=1}^T \mathbb{P}(\text{coupon } j \text{ in first } c \text{ selections}) \\ &= \sum_{j=1}^T \left(1 - \left(1 - \frac{1}{T}\right)^c\right) \\ &= \left(1 - \left(1 - \frac{1}{T}\right)^c\right) T, \end{aligned}$$

as desired. The asymptotic form follows immediately.

For part 2, we use McDiarmid's inequality [12], which characterizes the concentration of functions of independent random variables when the bounded difference property is satisfied. Write  $Y_1, Y_2, \dots, Y_c$  for the labels of the selected coupons, and  $W(c) = f(Y_1, Y_2, \dots, Y_c)$  for the number of distinct coupons. Note that here we have the bounded difference property, in that

$$|f(Y_1, \dots, Y_j, \dots, Y_c) - f(Y_1, \dots, \hat{Y}_j, \dots, Y_c)| \leq 1$$

for any  $j$ ,  $Y_1, \dots, Y_c$ , and  $\hat{Y}_j$ , since the largest difference we can make is swapping a distinct coupon  $Y_j$  for a non-distinct one  $\hat{Y}_j$ , or vice versa. McDiarmid's inequality [12] gives that

$$\mathbb{P}(|f(Y_1, \dots, Y_c) - \mathbb{E}f(Y_1, \dots, Y_c)| \geq \delta) \leq 2 \exp\left(-\frac{2\delta^2}{c}\right).$$

Setting  $\delta = \epsilon T$  and  $c = \alpha T$  gives the desired result; we crudely remove the factor of 2 from the exponent to account for the fact that we are considering deviations from the asymptotic mean instead of the true mean. ■

### IV. COMP

The COMP algorithm is due to Chan *et al.* [6]. It is based on the observation that any item in a negative test is definitely non-defective; COMP then declares all remaining items to be defective. Hence, COMP succeeds if and only if each of the  $N - K$  non-defectives appears in some negative test.

Chan *et al.* [6] show that COMP with Bernoulli tests succeeds with  $T > (1 + \epsilon)eK \ln N$  tests, for a rate of  $(1/e \ln 2)(1 - \theta) \approx 0.531(1 - \theta)$ , and Aldridge [1] showed that COMP can do no better than this with Bernoulli tests.

The following theorem reveals that constant weight columns improve the achievable rate by COMP.

*Theorem 1:* Consider group testing with a constant column weight design and the COMP algorithm. Write

$$T_{\text{COMP}}^* = \frac{1}{\ln 2} K \log_2 N.$$

Then, for any  $\epsilon > 0$ , with  $T \geq (1 + \epsilon)T_{\text{COMP}}^*$  tests, the error probability can be made arbitrarily small for  $N$  sufficiently large, while with  $T \leq (1 - \epsilon)T_{\text{COMP}}^*$  tests, the error probability is bounded away from 0. Hence, the maximum achievable rate is  $\ln 2(1 - \theta) \approx 0.693(1 - \theta)$ .

Note that this shows that COMP with a constant column weight design outperforms any algorithm with a Bernoulli design for large  $\theta$ , since it beats the rate in (1) for  $\theta > 1/(e(\ln 2)^2) \approx 0.766$ . Further, COMP improves the region where 'practical' algorithms work, in that it beats the best-known rate of a practical algorithm, previously given by the DD algorithm [3], for  $\theta < 1 - 1/(e \ln 2)^2 \approx 0.234$ .

*Proof:* We start by following [3, Remark 18]. Recall that we write  $L = \nu T/K$  for the weight of each column. We also write  $M$  for the number of positive tests. We know that COMP succeeds if and only if every non-defective item appears in a negative test. The probability that a given non-defective item appears in a negative test is 1 minus the probability it appears only in positive tests, which, conditional on  $M$ , is  $1 - (M/T)^L$ . Note that since our design has independent columns, the events that particular non-defective items appear in a negative test are conditionally independent given the list of positive tests; this also holds for Bernoulli designs, but not for constant row-and-column designs as studied by Wadayama [16].

Hence, given the number of positive tests, COMP has

$$\mathbb{P}(\text{success} | M) = \left(1 - \left(\frac{M}{T}\right)^L\right)^{N-K}. \quad (3)$$

We start with achievability. For any  $m^*$ , we have

$$\begin{aligned} \mathbb{P}(\text{success}) &= \mathbb{P}(M < m^*)\mathbb{P}(\text{success} | M < m^*) \\ &\quad + \mathbb{P}(M \geq m^*)\mathbb{P}(\text{success} | M \geq m^*) \\ &\geq \mathbb{P}(M \geq m^*)\mathbb{P}(\text{success} | M = m^*), \end{aligned}$$

since the success probability (3) is decreasing in  $M$ .

The number of positive tests  $M$  can be thought of as the number of distinct coupons collected from a population of  $T$  (all tests) by  $KL$  coupons:  $L$  for each column in a total of  $K$  columns corresponding to defective items. Hence, by Lemma 1, we have concentration of  $M$  about its mean  $\mathbb{E}M = (1 - e^{-\nu})T$ . Hence, setting  $m^* = (1 - e^{-\nu} - \epsilon)T$ , we have, for  $N$  sufficiently large,

$$\begin{aligned} \mathbb{P}(\text{success}) &\geq (1 - \epsilon)\mathbb{P}(\text{success} \mid M = (1 - e^{-\nu} - \epsilon)T) \\ &= (1 - \epsilon) \left( 1 - \left( \frac{(1 - e^{-\nu} - \epsilon)T}{T} \right)^L \right)^{N-K} \\ &= (1 - \epsilon) \left( 1 - (1 - e^{-\nu} - \epsilon)^{\nu T/K} \right)^{N-K}. \end{aligned}$$

It is easy to check by differentiating that  $(1 - e^{-\nu} - \epsilon)^\nu$  is maximised at  $e^{-\nu} = (1 - \epsilon)/2$ , which approaches  $\nu = \ln 2$  for small  $\epsilon$ . We therefore set  $\nu = \ln 2$ , yielding

$$\begin{aligned} \mathbb{P}(\text{success}) &\geq (1 - \epsilon) \left( 1 - \left( \frac{1}{2} - \epsilon \right)^{(\ln 2)T/K} \right)^{N-K} \\ &\geq (1 - \epsilon) \left( 1 - (N - K) \left( \frac{1}{2} - \epsilon \right)^{(\ln 2)T/K} \right) \\ &= (1 - \epsilon) \left( 1 - \exp \left( \frac{\ln 2}{K} T \ln(1/2 - \epsilon) + \ln(N - K) \right) \right). \end{aligned}$$

We see that the success probability can be made arbitrarily close to 1 by choosing  $\epsilon$  sufficiently small, provided that

$$T \geq -(1 + \epsilon) \frac{K \ln(N - K)}{\ln 2 \ln(1/2 - \epsilon)}.$$

The achievability result follows on noting that

$$-\frac{\ln(N - K)}{\ln(1/2 - \epsilon)} \rightarrow -\frac{\ln N}{\ln(1/2)} = \log_2 N$$

as  $N \rightarrow \infty$  and  $\epsilon \rightarrow 0$ , since  $K = o(N)$ .

We now turn to the converse. By a similar argument to the above, for any  $m^*$ , we have

$$\begin{aligned} \mathbb{P}(\text{success}) &= \mathbb{P}(M < m^*)\mathbb{P}(\text{success} \mid M < m^*) \\ &\quad + \mathbb{P}(M \geq m^*)\mathbb{P}(\text{success} \mid M \geq m^*) \\ &\leq \mathbb{P}(\text{success} \mid M = m^*) + \mathbb{P}(M \geq m^*). \end{aligned} \quad (4)$$

We now pick  $m^* = (1 - e^{-\nu} + \epsilon)T$ . By the concentration in Lemma 1, this gives that for any  $\epsilon$  and for  $N$  sufficiently large we have  $\mathbb{P}(M \geq m^*) \leq \epsilon$ . It remains to bound the first term of (4) by expanding out  $(1 - (1 - e^{-\nu} + \epsilon)^L)^{N-K}$  one term further, using the inclusion–exclusion formula, then bounding as before. We omit full details due to space constraints. ■

## V. ALGORITHM-INDEPENDENT CONVERSE

In this section, we give an upper bound on the rate of group testing with constant weight columns that holds for any detection algorithm and any choice of  $L$ .

The proof, which is based on the capacity converse for Bernoulli testing [1], depends on analysis of a particular algorithm called SSS, which was studied in detail by Aldridge, Baldassini and Johnson [3]. The SSS algorithm declares as its estimate of  $\mathcal{K}$  the *smallest satisfying set* (if a unique such set exists). A set  $\mathcal{L} \subseteq \{1, 2, \dots, N\}$  is *satisfying* for the design  $\mathbf{X}$  and outcomes  $\mathbf{y}$  if using design  $\mathbf{X}$  with true defective set  $\mathcal{L}$  would indeed give outcomes  $\mathbf{y}$ ; in other words, the estimate  $\mathcal{L}$  ‘satisfies’ the observations. The SSS algorithm chooses the satisfying set  $\mathcal{L}$  that minimises  $|\mathcal{L}|$ , if a unique such  $\mathcal{L}$  exists, and declares an error (say) otherwise.

**Theorem 2:** Consider group testing with a constant column weight design and any detection algorithm. Write

$$T^* = \max \left\{ K \log_2 \frac{N}{K}, \frac{1}{\ln 2} K \log_2 K \right\}. \quad (5)$$

Then, for any  $\epsilon > 0$ , with  $T \leq (1 - \epsilon)T^*$  the error probability is bounded away from 0.

Hence, using a constant weight column design, the capacity is bounded above by

$$\min \left\{ 1, \ln 2 \frac{1 - \theta}{\theta} \right\} \approx \min \left\{ 1, 0.693 \frac{1 - \theta}{\theta} \right\}.$$

*Proof:* We follow Aldridge’s proof of a similar result for Bernoulli testing [1]. As shown there, it suffices to bound the error probabilities of the COMP and SSS algorithms. We have bounded the rate of COMP in Theorem 1, and it easy to see that the COMP bound satisfies this theorem, since

$$\frac{1}{\ln 2} K \log_2 N \geq T^*,$$

where  $T^*$  is as in (5). Hence, it suffices to bound the error probability of the SSS algorithm, which we do below.

The first term in the maximum in (5) is the counting bound, which holds for arbitrary designs [5]. It remains to show the second term.

Following [3], the error probability of SSS is bounded by

$$\begin{aligned} \mathbb{P}(\text{error}) &\geq \sum_{j=1}^k (-1)^{j+1} \sum_{|\mathcal{J}|=j} \mathbb{P}(A_{\mathcal{J}}) \\ &\geq \sum_{|\mathcal{J}|=1} \mathbb{P}(A_{\mathcal{J}}) - \sum_{|\mathcal{J}|=2} \mathbb{P}(A_{\mathcal{J}}), \end{aligned}$$

where  $A_{\mathcal{J}}$  is the event that none of the  $T$  tests includes a defective item from  $\mathcal{J} \subseteq \mathcal{K}$ .

In coupon collector terms,  $A_{\mathcal{J}}$  is the event that the  $jL$  coupons of defective items from  $\mathcal{J}$  only hit the tests already hit by the  $(K - j)L$  coupons of the other defective items. Given  $S_{\mathcal{J}}$ , this number of ‘already hit’ tests, the probability that the other  $jL$  coupons only hit these tests is  $(S_{\mathcal{J}}/T)^{jL}$ .

Hence, we have, conditional on the  $S_{\mathcal{J}}$ s, that

$$\mathbb{P}(\text{error} \mid \{S_{\mathcal{J}}\}) \geq \sum_{|\mathcal{J}|=1} \left( \frac{S_{\mathcal{J}}}{T} \right)^{jL} - \sum_{|\mathcal{J}|=2} \left( \frac{S_{\mathcal{J}}}{T} \right)^{jL}. \quad (6)$$

We again parametrise  $L$  as  $L = \nu T/K$  for some (arbitrary)  $\nu > 0$ . From Lemma 1, the number of tests  $S_{\mathcal{J}}$  already hit

is exponentially concentrated about its mean, which behaves as  $\mathbb{E}S_{\mathcal{J}} \sim (1 - e^{-\nu(1-j/K)})T$ . Further, for  $j = 1, 2$ , this simplifies to  $\mathbb{E}S_{\mathcal{J}} \sim (1 - e^{-\nu})T$ .

We shall condition on the  $S_{\mathcal{J}}$  being suitably close to their means for  $|\mathcal{J}| = 1, 2$ . Specifically, we define the event

$$B = \left\{ \begin{array}{ll} S_{\mathcal{J}} \geq (1 - e^{-\nu} - \epsilon)T & \text{for all } |\mathcal{J}| = 1 \\ S_{\mathcal{J}} \leq (1 - e^{-\nu} + \epsilon)T & \text{for all } |\mathcal{J}| = 2 \end{array} \right\}.$$

Then by Lemma 1 and the union bound,

$$\mathbb{P}(B) \geq 1 - \left( K + \binom{K}{2} \right) 2 \exp \left( -\frac{\epsilon^2 KT}{\nu} \right) \geq 1 - \epsilon$$

for  $T$  sufficiently large.

Using a similar argument to Section IV, we have

$$\begin{aligned} \mathbb{P}(\text{error}) &= \mathbb{P}(B^c) \mathbb{P}(\text{error} | B^c) + \mathbb{P}(B) \mathbb{P}(\text{error} | B) \\ &\geq \mathbb{P}(B) \mathbb{P}(\text{error} | B) \\ &\geq (1 - \epsilon) \mathbb{P}(\text{error} | B) \end{aligned}$$

for  $N$  sufficiently large.

Combining the definition of the event  $B$  with (6), we have

$$\mathbb{P}(\text{error} | B)$$

$$\begin{aligned} &\geq \sum_{|\mathcal{J}|=1} \left( \frac{\mathbb{E}(S_{\mathcal{J}} | B)}{T} \right)^{jL} - \sum_{|\mathcal{J}|=2} \left( \frac{\mathbb{E}(S_{\mathcal{J}} | B)}{T} \right)^{jL} \\ &\geq \sum_{|\mathcal{J}|=1} \left( \frac{(1 - e^{-\nu} - \epsilon)T}{T} \right)^{jL} - \sum_{|\mathcal{J}|=2} \left( \frac{(1 - e^{-\nu} + \epsilon)T}{T} \right)^{jL} \\ &= K(1 - e^{-\nu} - \epsilon)^L - \frac{1}{2} K^2 (1 - e^{-\nu} + \epsilon)^{2L} \\ &= K(1 - e^{-\nu} - \epsilon)^{\nu T/K} - \frac{1}{2} K^2 (1 - e^{-\nu} + \epsilon)^{2\nu T/K} \\ &= K(1 - e^{-\nu} - \epsilon)^{\nu T/K} \left( 1 - \frac{K}{2} \left( \frac{(1 - e^{-\nu} + \epsilon)^2}{1 - e^{-\nu} - \epsilon} \right)^{\nu T/K} \right). \end{aligned}$$

For  $\epsilon$  sufficiently small, this is, like before, minimized arbitrarily close to  $\nu = \ln 2$ . Also for  $\epsilon$  sufficiently small,

$$\frac{(1 - e^{-\nu} + \epsilon)^2}{1 - e^{-\nu} - \epsilon}$$

is arbitrarily close to  $1 - e^{-\nu}$ . Hence, using  $\nu = \ln 2 + o(1)$ , we have, as  $\epsilon \rightarrow 0$ ,

$$\begin{aligned} \mathbb{P}(\text{error} | B) &\geq (1 - o(1)) K(1 - e^{-\ln 2})^{(\ln 2)T/K} \\ &\quad \times \left( 1 - \frac{1}{2} K(1 - e^{-\ln 2})^{(\ln 2)T/K} \right), \end{aligned}$$

where we have absorbed all  $o(1)$ s into the first term. But

$$\begin{aligned} K(1 - e^{-\ln 2})^{(\ln 2)T/K} &= K \left( \frac{1}{2} \right)^{(\ln 2)T/K} \\ &= 2^{\log_2 K - (\ln 2)T/K}, \end{aligned}$$

and since the error probability is decreasing in  $T$ , we have for

$$T \leq (1 - \epsilon) \frac{1}{\ln 2} K \log_2 K.$$

that the error probability for SSS is bounded away from 0. ■

## VI. CONCLUSIONS AND FUTURE WORK

We have shown that constant column weight matrix designs, instead of Bernoulli designs, can improve the rate that can be proved for the simple, yet order optimal, COMP algorithm. Further, we have shown an algorithm-independent converse, giving an upper bound on the rate which can be achieved by any algorithm under this design. We conjecture that this converse is sharp, in the sense that this performance can asymptotically be achieved by the SSS algorithm, and even by the simpler DD algorithm for certain parameter values. We hope to adapt the techniques of [3], [14] to resolve this conjecture in future work.

## ACKNOWLEDGMENTS

M. Aldridge was supported by the Heilbronn Institute for Mathematical Research. J. Scarlett was supported by the ‘EPFL fellows’ programme (Horizon2020 grant 665667).

## REFERENCES

- [1] M. Aldridge. The capacity of Bernoulli nonadaptive group testing. See [arxiv:1511.05201](https://arxiv.org/abs/1511.05201), 2015.
- [2] M. Aldridge, L. Baldassini, and K. Gunderson. Almost separable matrices. *Journal of Combinatorial Optimization (to appear)*, 2015. doi:10.1007/s10878-015-9951-1.
- [3] M. Aldridge, L. Baldassini, and O. T. Johnson. Group testing algorithms: bounds and simulations. *IEEE Trans. Inform. Theory*, 60(6):3671–3687, 2014.
- [4] G. Atia and V. Saligrama. Boolean compressed sensing and noisy group testing. *IEEE Trans. Inform. Theory*, 58(3):1880–1901, March 2012.
- [5] L. Baldassini, O. T. Johnson, and M. Aldridge. The capacity of adaptive group testing. In *Proceedings of the 2013 IEEE International Symposium on Information Theory, July 2013*, pages 2676–2680, 2013.
- [6] C. L. Chan, P. H. Che, S. Jaggi, and V. Saligrama. Non-adaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms. In *Proceedings of the 49th Allerton Conference on Communication, Control, and Computing*, pages 1832–1839, 2011.
- [7] R. Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, pages 436–440, 1943.
- [8] D. Du and F. Hwang. *Combinatorial Group Testing and Its Applications*. Series on Applied Mathematics. World Scientific, 1993.
- [9] F. K. Hwang. A method for detecting all defective members in a population by group testing. *Journal of the American Statistical Association*, 67(339):605–608, 1972.
- [10] O. T. Johnson. Strong converses for group testing in the finite blocklength regime. In submission, see: [arxiv:1509.06188](https://arxiv.org/abs/1509.06188), 2015.
- [11] W. Kautz and R. Singleton. Nonrandom binary superimposed codes. *IEEE Trans. Inform Theory* 10(4): 363–377, 1964
- [12] C. McDiarmid. On the Method of Bounded Differences. *Surveys in Combinatorics* 141:148–188, 1989
- [13] M. Mézard, M. Tarzia, and C. Toninelli. Group testing with random pools: Phase transitions and optimal strategy. *Journal of Statistical Physics*, 131(5):783–801, 2008.
- [14] J. Scarlett and V. Cevher. Limits on support recovery with probabilistic models: An information-theoretic framework. [http://arxiv.org/abs/1501.07440](https://arxiv.org/abs/1501.07440), 2015.
- [15] J. Scarlett and V. Cevher. Phase transitions in group testing. In *Proceedings of the 27th ACM-SIAM Symposium on Discrete Algorithms*, pages 40–53, 2016.
- [16] T. Wadayama. An analysis on non-adaptive group testing based on sparse pooling graphs. In *Proceedings of the 2013 IEEE International Symposium on Information Theory*, pages 2681–2685, 2013.